

Methodology:-

- json is converted into dataframe
- dataframe is pre-processed to validate data, cleaning up, remove all useless columns.
- If date of any row is invalid then that row is removed.
- Rows whose dates are outside start_date and end_date are removed.
- Rows are sorted by dates.
- 3 dataframes are created from the dataframe above.(For confirmed, recovered, deceased) and 'status' column is removed from those 3.
- Count of confirmed, recovered, deceased cases in all states is assumed to be an **integer** and if found not integer then it is **replaced by zero** so that we can perform integer operations on that.

Assumptions :-

- states = ['ap', 'ar', 'as', 'br', 'ct', 'ga', 'gj',
'hp', 'hr', 'jh', 'ka', 'kl', 'mh', 'ml', 'mn',
'mp', 'mz', 'nl', 'or', 'pb', 'rj', 'sk', 'tg',
'tn', 'tr', 'up', 'ut', 'wb']
- uts = ['an', 'ch', 'dd', 'dl', 'dn', 'jk', 'la', 'ld', 'py']
- **Delhi** is a Union territory.
- 'un' column is removed from dataset.
- in Q1_1, '**tt**' is not used, all states and Uts are considered.
- start_date and end_date format – **YYYY-MM-DD** and dates are assumed to be **valid**, otherwise proper message will be displayed.
- In Q_3, **start_date** is considered as origin.
- If some state is missing in a row, then its value is assumed to be zero.

Output :
2018101 and 2018261

Q1_1 :-

confirmed_count: 4110214 recovered_count: 3177666 deceased_count: 70095

Q1_2 :-

confirmed_count: 188193 recovered_count: 163785 deceased_count: 4538

Q1_3 :-

confirmed_count: 1072055 recovered_count: 800359 deceased_count: 30813

Q1_4 :-

Confirmed :-
Highest affected State is: ['mh']
Highest affected State count is: 883862

Recovered :-
Highest affected State is: ['mh']
Highest affected State count is: 636574

Deceased :-
Highest affected State is: ['mh']
Highest affected State count is: 26275

Q1_5 :-

Confirmed :-
Lowest affected State is: ['mz']
Lowest affected State count is: 1062

Recovered :-
Lowest affected State is: ['mz']
Lowest affected State count is: 713

Deceased :-
Lowest affected State is: ['mz']
Lowest affected State count is: 0

Q1_6 :-

Confirmed :-
Day: 2020-06-23
Count: 3947

Recovered :-
Day: 2020-06-20
Count: 7725

Deceased :-
Day: 2020-06-16
Count: 437

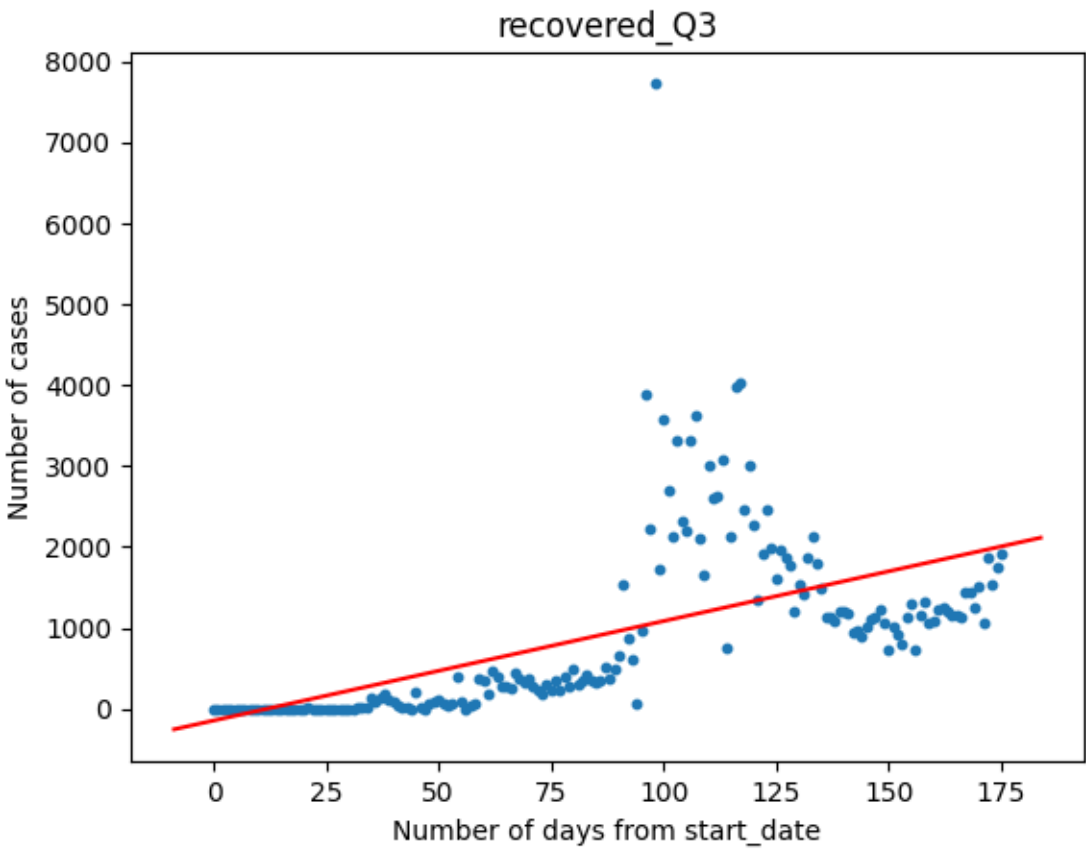
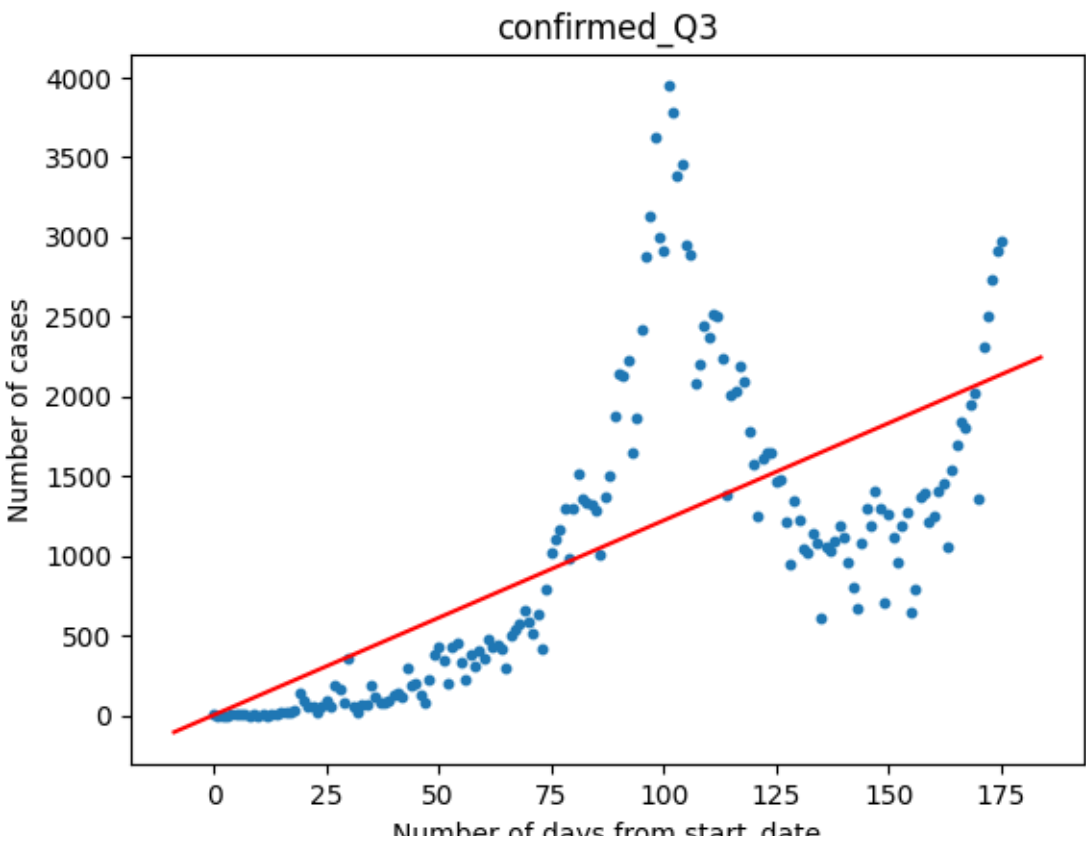
Q1_7 :-

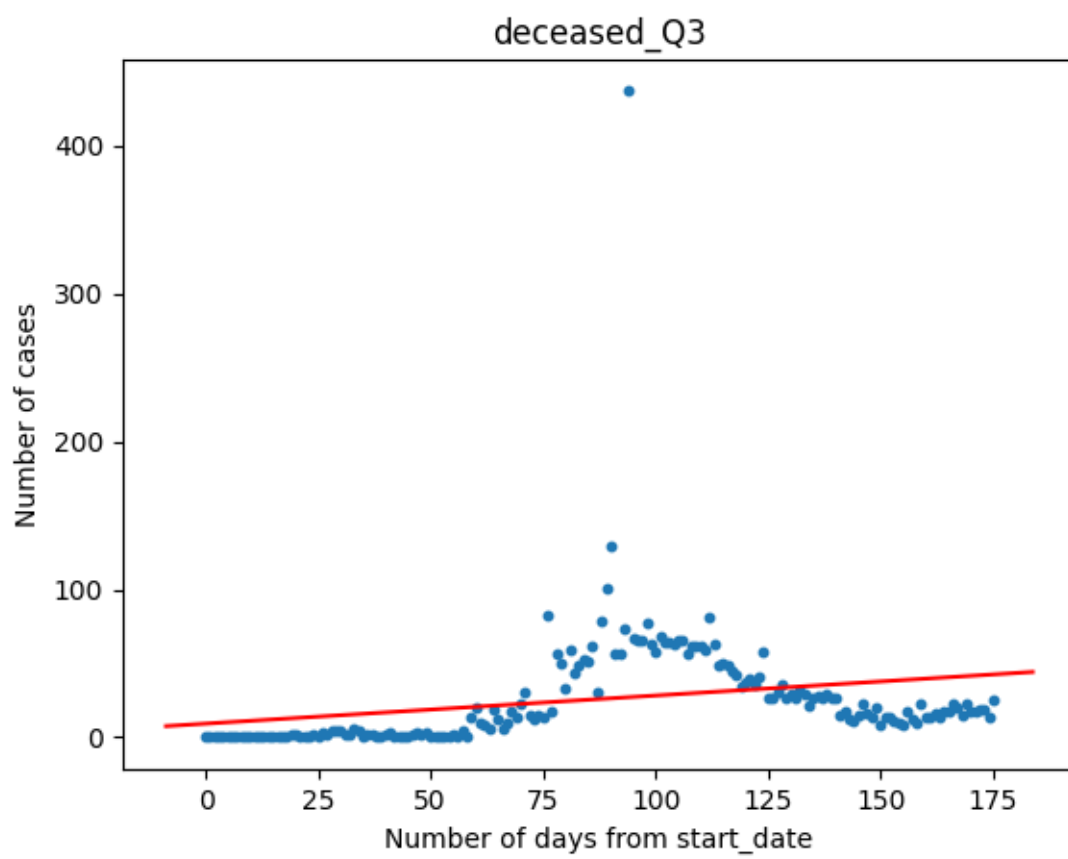
State	Active_Cases
ap	100880
ar	1525
as	28404
br	16735
ct	22320
ga	4945
gj	16266
hp	2023
hr	14912
jh	14980
ka	100224
kl	21867
mh	221013
ml	1374
mn	1872
mp	15687
mz	349
nl	701
or	25856
pb	15870
rj	14996
sk	561
tg	32405
tn	51580
tr	5905
up	59963
ut	7649
wb	23390

Q3 :-

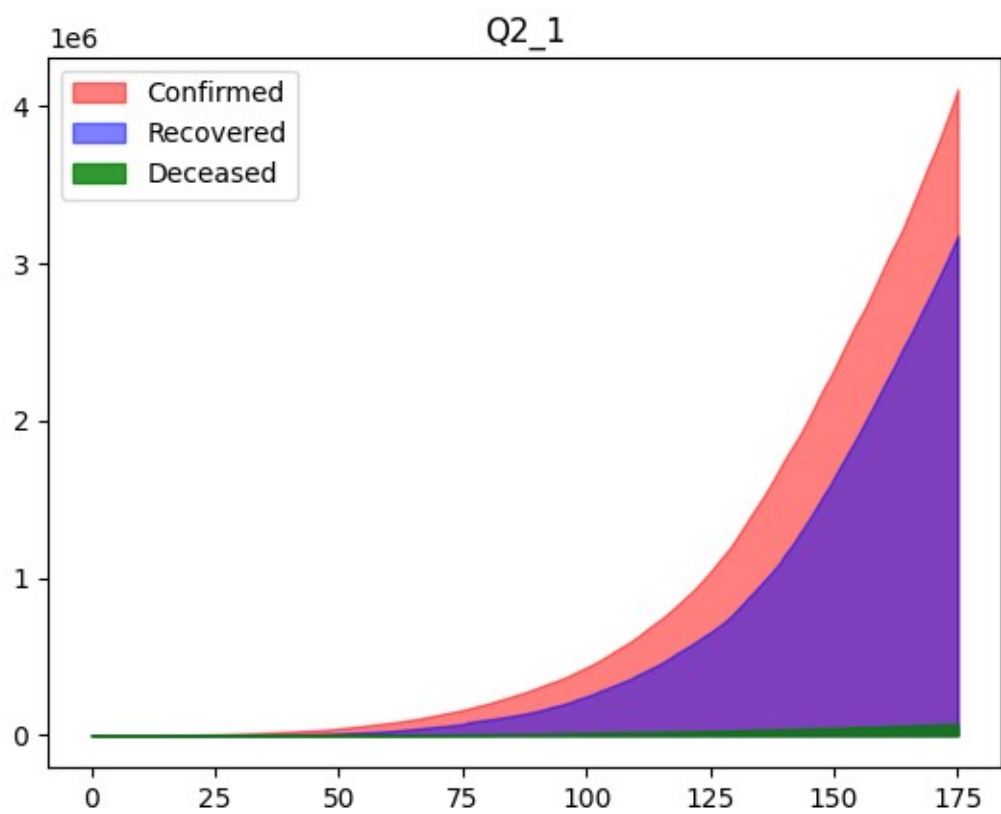
0.5298536209553159 12.2142692053709 -146.1371340523883 12.305528285274049
9.138674884437597 0.19023332599603787

Q3 plots

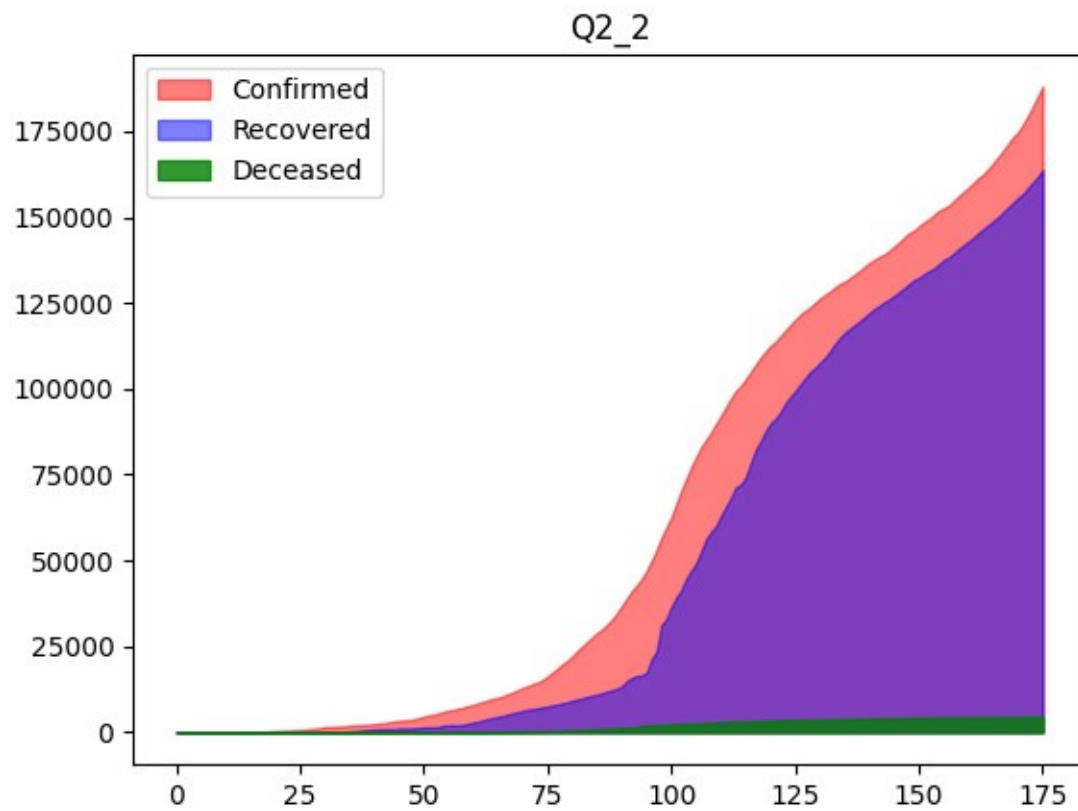




Q2_1 plots :



Q2_2 plot:



Q2_3:

